White Paper | Public Sector

Simplifying RFP Evaluations through Human and GenAI Collaboration

GenAI and human expertise improve procurement decision-making and RFP evaluations.

Authors

Dr. Darren W Pulsipher

Abstract

RFP evaluations can be costly and inconsistent. This white paper discusses how integrating Generative AI (GenAI) with human expertise enhances efficiency, accuracy, and decision-making. Organizations can streamline workflows, reduce biases, and make better procurement decisions by automating data processing and applying human oversight. Explore the benefits, best practices, and real-world applications of Human-Augmented AI in transforming RFP evaluations and optimizing procurement.

Table of Contents

Overview

The increasing complexity and volume of Requests for Proposals (RFPs), Requests for Information (RFIs), and contracts present significant challenges for organizations. With the widespread adoption of large language models (LLMs), vendors can now generate highly detailed and comprehensive responses, increasing the workload for evaluators. Traditional evaluation methods, which rely heavily on manual review processes, struggle to keep pace with this influx of information—handling sensitive procurement data results in inefficiencies, inconsistencies, and potential security risks.

Private GenAI offers a revolutionary approach to processing and evaluating requests for proposals (RFPs), requests for information (RFIs), and contracts. By ensuring data privacy and security in today's digital landscape, Private GenAI empowers organizations to conduct these activities securely on-premises. Unlike cloud-based AI solutions, which can expose sensitive information to external vulnerabilities, Private GenAI guarantees that every piece of data remains within the confines of an organization's controlled infrastructure. This instills a sense of security and confidence, putting organizations in control of their procurement processes.

March 17, 2025

While AI technology, such as Private GenAI, can undoubtedly enhance evaluations, the role of human analysts in the procurement evaluation process remains crucial. Their unique abilities, such as critical thinking and nuanced judgment, cannot be fully replicated by AI systems. By leveraging AI-powered scorecarding techniques, we can streamline evaluations and allow analysts to focus on strategic aspects, ensuring they feel valued and integral to the process.

Evaluating RFPs, RFIs, and contracts is increasingly challenging due to the labor-intensive nature of manual assessments, mainly as vendors utilize AI-generated responses. The growing complexity and volume of submissions have rendered traditional evaluation methods impractical. This is compounded by the need for compliance with legal and ethical standards, as skilled analysts provide oversight to ensure alignment with regulatory requirements and adaptability to unique situations that AI may not effectively address.

Problem Statement

Evaluating RFPs, RFIs, and contracts is traditionally labor-intensive, requiring analysts' teams to assess submissions manually. As vendors increasingly use AI-generated responses, these documents' sheer volume and complexity have made manual evaluation impractical. Several key challenges hinder practical procurement evaluation:

Complexity and Volume

The sheer volume of RFP, RFI, and contract submissions has expanded exponentially, creating severe bottlenecks for procurement and evaluation teams. Each submission contains extensive technical, operational, and legal details that require scrutiny to ensure compliance and value alignment with organizational objectives. This necessitates countless hours of manual analysis to extract relevant insights, validate the accuracy of details, and compare against competitors. The growing complexity often results in delays and oversights, which can have financial and operational repercussions for the organization.

Moreover, as vendors increasingly utilize AI tools to generate detailed and polished submissions, the workload for evaluators grows exponentially. When relying solely on human resources, identifying nuances, verifying facts, and standardizing assessments in lengthy vendor responses become an impractical challenge. These inefficiencies compound over time, leaving organizations struggling to meet deadlines, undermining their confidence in procurement processes, and exposing them to risks associated with poorly reviewed contracts or proposals.

Subjectivity and Inconsistency

When driven solely by human review, procurement evaluations are vulnerable to subjective interpretation and inconsistencies. Evaluators often bring individual biases or differing levels of expertise into the process, affecting how scoring criteria are interpreted and applied. Consequently, identical proposals may receive varying scores depending on the evaluators' perspectives or familiarity with specific evaluation requirements. This lack of standardization affects the credibility of results and opens the door to vendor disputes, eroding trust in the evaluation process.

Inconsistent evaluations also hinder strategic decision-making. Disorganized results often obscure patterns and trends that should inform procurement strategies. A consistent scoring model is essential, but achieving it manually becomes increasingly unfeasible in light of documentation's growing volume and complexity. The absence of standardization impairs the ability to extract actionable intelligence, pushing organizations to seek innovative solutions that foster reliability and consistency.

Data Security Risks

Cloud-based AI solutions introduce significant concerns regarding data privacy and security. Traditional evaluation platforms often require sensitive procurement materials, such as financial agreements, product specifications, and legal clauses, to be uploaded to external servers for processing. This creates exposure to potentially unauthorized access, both from external threats (e.g., hackers) and internal vulnerabilities within third-party service providers. Regulatory compliance further compounds the issue, as the handling of procurement data must adhere to specific industry and geographic norms, such as GDPR or HIPAA, depending on the organization's domain.

The risks associated with cloud dependency go beyond merely unauthorized access. Organizations risk losing control over their proprietary data, as many public AI models retain training histories that could inadvertently include sensitive information. These concerns escalate compliance anxiety and deter organizations from adopting AI solutions altogether. An on-premises model—such as Private GenAI—provides a clear resolution to this challenge by ensuring that all data processing occurs within the organization's secure infrastructure, safeguarding sensitive materials from external exposure.

Hallucinations in AI Models

The reliability of public generative AI models is compromised by phenomena such as 'hallucinations,' wherein the AI generates outputs that are factually inaccurate, misleading, or entirely fabricated. These issues are particularly concerning in procurement evaluations, where accurate interpretation of technical and legal documents is critical. Misrepresentations or incorrect assessments [1] AI could lead to flawed decisions, such as awarding contracts to incapable vendors or misjudging proposals' compliance. It's important to note that 'hallucinations' in AI models are not intentional but rather a result of the model's training data and the complexity of the task.

Unchecked reliance on public AI tools raises questions about accountability and quality assurance in procurement processes. For example, hallucinated responses, which are AI-generated outputs that are factually inaccurate, misleading, or entirely fabricated, could propagate errors across evaluative reports, undermining the trust of stakeholders and decision-makers. Private GenAI integrates AI's computational strength with human expertise, ensuring every AI-suggested outcome undergoes careful review by evaluators. This hybrid approach enhances accuracy while maintaining confidence in automated evaluation processes, countering the limitations of traditional machine-based models.

Technology Overview

Private GenAI and Localized AI Processing

Private GenAI represents a strategic shift in artificial intelligence deployment, focusing on enhancing security, customization, and performance by operating exclusively within an organization's secure infrastructure. Unlike traditional cloud-based AI solutions, Private GenAI leverages on-premises resources such as departmental LLM clusters or AI-enabled workstations (AI PCs) with advanced hardware like Core Ultra chips or Xeon-based servers. This localized approach helps organizations address critical challenges in data privacy, regulatory compliance, and accuracy.

The primary advantage of Private GenAI lies in its capability to confine sensitive data within the organization's ecosystem. This approach prevents exposure to external risks, such as unauthorized access or potential data breaches, when transmitting information to third-party platforms. Furthermore, organizations gain independence from third-party cloud AI vendors, allowing them to tailor models to their specific needs, such as optimizing models for domain-specific procurement evaluations. Fine-tuning AI outputs for accuracy and precision further empowers automated processes like scorecards, resulting in faster and more accurate assessment of RFPs, RFIs, and contracts. Ultimately, Private GenAI enhances operational efficiency without compromising security or performance, making it a transformative solution for enterprises.

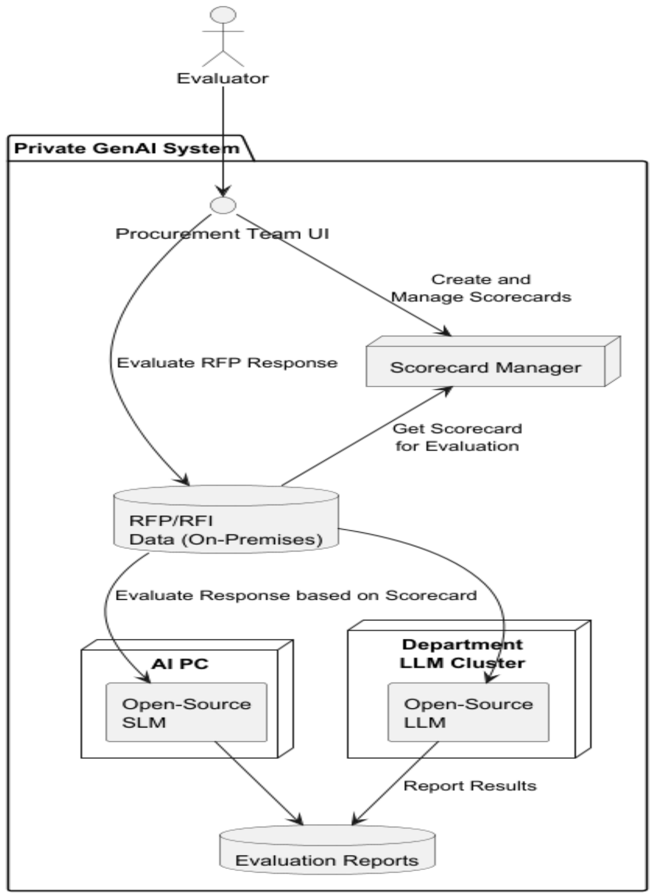

Figure 1 High-level Technology

AI PC with Open-Source LLMs/SLMs

An AI PC is a powerful, purpose-built workstation designed to run large language models (LLMs) or specialized language models (SLMs) in a secure, local environment. These systems utilize powerful processing units, including AI-enhanced CPUs and GPUs, to support robust on-premises AI applications. In the context of Private GenAI, AI PCs enable organizations to process and evaluate RFPs, RFIs, and contracts on-premises, ensuring that sensitive information remains within the organization's controlled infrastructure. By incorporating open-source models like Llama, Mistral, or Mixtral, AI PCs allow organizations to integrate cutting-edge capabilities without the constraints or risks associated with proprietary AI solutions. [2].

The use of AI PCs yields significant benefits for procurement evaluations. Data remains on-premises throughout the entire lifecycle, eliminating the vulnerability introduced by cloud-based systems. Moreover, open-source LLMs provide a customizable foundation for creating AI applications tailored to organizational requirements and internal evaluation frameworks. For example, procurement scoring methodologies can be embedded directly into these models, ensuring alignment with internal processes and compliance standards. In addition, the performance of an AI PC is optimized for real-time processing of complex data, significantly reducing evaluation time while maintaining accuracy. These localized systems' flexibility, security, and speed ensure they are well-suited for managing the increasing complexity of large-scale procurement processes.

Security and Compliance Considerations

Integrating Private GenAI and AI PCs mitigates security risks and ensures compliance with ever-evolving regulatory requirements. By operating entirely on-premises and within the organization’s controlled infrastructure, Private GenAI provides high protection for confidential procurement data. Only authorized personnel can access sensitive information, significantly reducing the risk of internal leaks or unauthorized usage. These access controls align closely with industry best practices and safeguard data against potential threats.

Additionally, deploying localized AI systems ensures adherence to regional, industry, and corporate compliance standards. Regulatory frameworks such as GDPR, HIPAA, or industry-specific requirements demand stringent data handling practices, and Private GenAI facilitates seamless compliance by confining data processing to a secure local environment. Beyond regulatory compliance, localized AI processing eliminates vendor lock-in risks, offering flexibility and cost control. Organizations can upgrade, modify, or switch models as operational needs evolve, avoiding the restrictive licensing terms imposed by proprietary cloud platforms. Together, these capabilities ensure secure, compliant AI deployment tailored to organizational requirements.

Scorecarding for Evaluations

What is Scorecarding?

Scorecarding is an assessment methodology that brings structure and consistency to evaluating RFPs, RFIs, and similar procurement documents. [1]. Organizations can apply a fair, transparent, and repeatable process to identify the most qualified vendors by assigning quantifiable scores to responses using predefined criteria. Scorecarding is particularly valuable in complex procurement scenarios where subjective judgments may otherwise dominate the decision-making process. The approach ensures a systematic method to compare and contrast multiple submissions, reducing evaluation discrepancies objectively.

The primary benefits of scorecarding lie in its ability to standardize evaluations and reduce subjectivity [2]. Standardization ensures that evaluators across different teams or departments adhere to a shared framework, improving consistency in results. By replacing free-form qualitative assessments with measurable metrics, scorecarding minimizes the influence of individual biases and varying levels of expertise among evaluators. Additionally, this structured process enables procurement teams to make data-driven decisions, prioritizing high-quality responses efficiently and ensuring procurement outcomes align with organizational goals.

The Role of AI in Scorecarding

Artificial intelligence significantly enhances the traditional scorecard process by automating and refining key evaluation stages. AI-powered systems can analyze complex and voluminous RFP/RFI responses, extracting relevant information quickly and precisely. This automation minimizes manual burdens, enabling procurement teams to focus on reviewing outputs and higher-value tasks. AI excels at identifying elements that align with predefined evaluation criteria and categorizing and prioritizing them for scoring.

Beyond preliminary scoring, AI also provides justifications for its scoring decisions. These explanations are invaluable in validating results, offering analysts clear visibility into the rationale behind scores, and enabling confidence in automated outputs. [3]. AI ensures assessment accuracy, relevance, and depth by referencing evaluation benchmarks and matching informatively extracted content from submissions. This accelerates the decision for procurement teams and fortifies the scorecarding process with data-driven insights, fostering informed decision-making while reducing oversight risks.

Analyst-AI Collaboration

In the scorecard process, AI augments human expertise rather than replaces it. Analysts retain control over the evaluation process, ensuring domain knowledge, organizational priorities, and regulatory considerations guide the final judgments. AI-generated scores provide a starting point, but analysts bring critical insights to validate, refine, and adjust these outputs wherever necessary. This collaborative approach ensures that human expertise remains at the core of decision-making while streamlining workflows through automation.

The partnership between analysts and AI fosters transparency and accountability. Analysts can review AI annotations and references to verify the accuracy of extracted content, ensuring that critical procurement requirements are met. The AI's ability to provide clear justifications for its scores further aids the review process, enabling evaluators to identify gaps, inconsistencies, or oversights. By combining the speed and objectivity of AI with the nuanced judgment of human evaluators, this collaboration strengthens confidence in procurement decisions, ensuring they are both efficient and reliable.

AI-Assisted Evaluation Process

An AI-assisted evaluation workflow provides a structured and efficient framework for assessing RFP/RFI submissions. By combining AI and human expertise, organizations can expedite the evaluation process, improve accuracy, and maintain transparency. Below is a step-by-step workflow example demonstrating how organizations can implement AI in their procurement evaluations while ensuring clarity and control at every stage.

Step 1: Submission Processing

The workflow begins with the ingestion and pre-processing of RFP/RFI submissions. Documents are uploaded and processed in a secure, on-premises environment using an AI-enabled system like an AI PC. During this phase, unstructured data in various formats—PDFs, Word documents, and spreadsheets—is converted into a structured, machine-readable format. The AI system automatically performs text extraction, document classification, and segmentation tasks, ensuring that each submission section (e.g., technical details, pricing information, compliance responses) is adequately identified and organized.

This automation saves significant time by eliminating the need for manual processing. For example, an organization evaluating ten vendor proposals typically spends hours categorizing responses. With AI, this process can be completed in mere minutes, allowing evaluators to move to the next phase of the workflow quickly. Additionally, the system flags missing or incomplete sections of submissions, providing visibility into potential gaps early in the process.

Step 2: Criteria Matching

Once the submissions are processed, AI applies the organization's predefined scoring rules to evaluate the content against specific procurement criteria. These criteria could include technical compliance, pricing competitiveness, delivery timelines, or sustainability metrics. The AI system analyzes the details provided in the vendor responses and cross-references them with the scoring standards set by the organization.

For instance, if one evaluation criterion is "proven experience in large-scale system implementation," the AI will extract statements or evidence from the vendor's proposal indicating relevant experience and score the response accordingly. At this stage, the system also highlights how each response aligns with the evaluation criteria, making it easy for evaluators to understand why a particular score has been awarded. This automated analysis ensures objectivity by applying the same benchmarks uniformly to all submissions and accelerates the evaluation process significantly.

Step 3: AI-Generated Scorecard

After matching the criteria, the AI system generates an initial scorecard for each submission. This scorecard provides detailed scores for each evaluation criterion and supporting references from the original vendor response. For example, if a vendor claims compliance with a technical standard, the scorecard will include excerpts from their proposal as evidence for the score.

The scorecard serves as the foundation for the subsequent analyst review phase. It presents a clear and easily digestible view of how each submission performs against predefined benchmarks. It enables evaluators to identify strengths and weaknesses without sifting through hundreds of documentation pages. By providing justifications for each score, the scorecard ensures transparency, helping evaluators trust the AI's outputs and focus on refining the results where necessary.

Step 4: Analyst Review

While AI significantly speeds up evaluation and reduces manual effort, human oversight remains critical. During this step, analysts review the AI-generated scorecards to validate the scores, ensure relevance, and account for nuanced factors that the AI may not have captured. For example, if a vendor indicates compliance with a standard but provides insufficient evidence, an evaluator can adjust the score to reflect this limitation.

In addition to adjusting scores, evaluators verify the AI's references to ensure accuracy and completeness. Analysts may also document additional observations or involve subject matter experts for specific criteria that require more profound technical knowledge. This collaborative approach ensures the final evaluations are accurate and aligned with the organization's strategic goals and procurement priorities.

Step 5: Final Decision

Once the analyst review is complete, the procurement team consolidates the validated scores to finalize rankings. Based on the final scores, vendors are ranked according to their suitability for the contract. For example, a vendor scoring the highest on technical capability, cost-effectiveness, and compliance may be prioritized for contract award. Additionally, the team can generate comprehensive evaluation reports for internal stakeholders, ensuring decisions remain well-documented and justifiable.

This hybrid approach, where AI accelerates the initial stages and human analysts refine and validate outputs, enhances transparency and consistency in the decision-making process. Organizations can make procurement decisions confidently and efficiently by significantly reducing evaluation time while maintaining precision and accountability.

Example Application

Let's consider an organization evaluating bids for a multimillion-dollar IT system upgrade. Ten vendors submit proposals with details ranging from technical compliance and implementation timelines to pricing and support services. Using an AI-assisted scorecard, the following workflow is applied:

Submission Processing: The AI ingests the submissions, extracts information into structured datasets, and identifies sections such as "Experience," "Pricing," and "Technical Specifications."

Criteria Matching: The AI evaluates how well each submission aligns with key criteria such as "proven project experience," scoring vendors based on the number and quality of the examples they provide.

AI-Generated Scorecards: Vendor A, for example, receives a preliminary score of 9/10 for project experience, justified by references to five previous large-scale implementations cited in their proposal.

Analyst Review: Analysts verify Vendor A's claims, find one reference misaligned, and adjust the score to 8/10. Additionally, experts weigh in on the technical solution's feasibility, making minor refinements to other areas of evaluation.

Final Decision: Vendor scores are consolidated, and the procurement team ranks Vendor A as the top choice. The contract is awarded after ensuring alignment with budgetary constraints and organizational goals.

This workflow illustrates how organizations can blend AI's efficiency with human expertise to create a repeatable, scalable evaluation process suited to modern procurement challenges.

Implementation Strategy

Setting Up an AI PC for RFP/RFI Evaluation

Deploying an AI PC is key to modernizing procurement evaluations by enabling secure and efficient on-premises processing. AI PCs are purpose-built systems with advanced hardware, such as AI-enhanced CPUs and GPUs, that are optimized for handling large language model (LLM) inference with high performance.[6]. These workstations provide the computational power to efficiently process lengthy and complex RFP/RFI documents, extracting key insights and generating structured evaluations in real-time.

In addition to hardware, robust software capabilities are critical to the AI PC setup. Open-source AI models, such as those supporting Retrieval-Augmented Generation (RAG) techniques, are particularly suited for these tasks. RAG enables models to retrieve and utilize data from a local knowledge base to ensure contextual accuracy and relevance [5]. Organizations must also integrate local databases of procurement criteria, past vendor responses, and regulatory frameworks to enhance the model's contextual understanding further. This blend of hardware and software ensures precise and secure RFP/RFI evaluations within the organization's infrastructure.

Training and Customization

Traditional fine-tuning and training AI models to meet domain-specific requirements can be time-consuming and cost-prohibitive, particularly for organizations with unique procurement evaluation needs. However, leveraging alternative approaches such as Retrieval-Augmented Generation (RAG), advanced prompt engineering, and scorecarding [6] Can address these challenges while delivering highly accurate and efficient results. These strategies eliminate the need for extensive fine-tuning by leveraging existing knowledge bases, structuring queries effectively, and automating evaluation processes.

RAG provides a powerful alternative to fine-tuning by augmenting general-purpose AI models with domain-specific data retrieved from a local or pre-curated knowledge base [7]. Instead of embedding organizational requirements into the model, RAG dynamically retrieves relevant information during inference, such as historical RFP responses, guidelines, or scoring rules. For example, when evaluating vendor submissions, the AI retrieves and incorporates policy documents or benchmarks to generate contextually relevant and precise outputs without the overhead of re-training the model. By dynamizing organizational knowledge, RAG significantly reduces implementation costs and the lead time for deploying customized AI solutions.

Prompt engineering further minimizes the need for customized training by enabling organizations to guide general-purpose AI models through carefully designed, task-specific prompts. These prompts act as templates that structure how the AI interacts with input data, ensuring it focuses on the relevant aspects of an RFP or RFI response. For instance, prompts can direct the AI to extract key metrics like compliance statements, technical specifications, or sustainability commitments from submissions and align them with specific evaluation criteria. This method enhances model performance in a targeted way without requiring modifications to the underlying AI, making it a cost-effective alternative to traditional fine-tuning.

Combined with RAG and prompt engineering, the scorecard further streamlines the evaluation process by applying a structured framework to vendor assessments. The AI uses RAG techniques to retrieve context for each evaluation criterion, while prompt engineering ensures it applies the correct logic to generate scores. These scores are then presented through an AI-assigned scorecard, highlighting how vendor responses align with evaluation benchmarks. This hybrid approach achieves the precision and domain specificity of fine-tuned models without the extensive investment of time and resources. It is an ideal solution for procurement teams looking to modernize their RFP/RFI workflows efficiently.

By combining RAG, prompt engineering, and scorecarding, organizations can optimize AI deployments for procurement evaluations without the cost and time burdens of traditional model fine-tuning, enabling faster, more accurate, and more scalable decision-making processes.

Integration with Existing Procurement Workflows

A successful implementation strategy ensures that Private GenAI solutions seamlessly integrate into procurement workflows. AI PCs can be connected to document management and procurement platforms to avoid disruptions and to ensure smooth data exchange. For instance, incoming RFPs or RFIs can be ingested directly from document portals, and evaluation results can be automatically pushed back into procurement systems. This integration significantly reduces manual handovers and paperwork, enabling a more cohesive workflow.

API-driven integration further extends the functionality of the AI-assisted evaluation system. By utilizing APIs, the AI can communicate with other enterprise systems—for example, pulling evaluation criteria from procurement software, accessing vendor performance history from CRM systems, or feeding validated vendor rankings directly into ERP platforms. This interoperability allows organizations to maintain a unified data ecosystem and enhances operational efficiency by eliminating redundancies and fostering real-time insights across departments. By embedding Private GenAI into the broader procurement pipeline, organizations can make their workflows faster, more scalable, and highly adaptive.

Adoption and Best Practices

Organizations must take a strategic and structured approach to integrate Private GenAI into procurement evaluations successfully. While AI offers tremendous potential to enhance efficiency, accuracy, and transparency in decision-making processes, its adoption requires careful planning and balanced implementation. Organizations can create sustainable and effective AI-powered workflows tailored to their unique procurement needs by focusing on clear guidelines, human collaboration, and iterative improvements.

One of the first steps in adoption is the creation of well-documented guidelines for AI-assisted evaluations. These guidelines should define the scope of AI involvement, outline evaluation criteria, and establish standards to ensure consistent and fair assessments. For example, organizations should determine which aspects of the procurement process will be fully automated—such as criteria matching and preliminary scoring—and which will require human intervention, such as final score validation and nuanced decision-making. By documenting these processes, organizations can ensure transparency and alignment across teams while setting clear expectations for the role of AI in their workflows.

Maintaining a balance between automation and human oversight is critical throughout the adoption process. While AI can handle repetitive, data-heavy tasks such as information extraction and initial scoring, human evaluators must remain in control of crucial decisions to ensure accuracy and accountability. Analysts should validate AI-generated outputs, refine scores where needed, and consider AI results within the broader organizational context. This collaborative approach ensures high-quality evaluations and builds trust among stakeholders by emphasizing the human element in sensitive procurement decisions.

Finally, continuous refinement of AI systems is essential for long-term success. Organizations should collect feedback from evaluators and use these insights to improve the AI's performance incrementally. Regular updates to scoring rules, prompts, or retrieval techniques based on changing procurement policies or industry trends ensure the system stays relevant and adaptable. Over time, this iterative process will help build a highly optimized AI platform capable of handling increasingly complex procurement challenges with greater precision.

As AI-driven procurement evolves, future advancements in Private GenAI will further improve efficiency, accuracy, and security. Capabilities such as more advanced contextual analysis, better integration with enterprise systems, and stricter compliance with regulatory frameworks will enhance the value AI delivers to procurement teams. By adopting best practices today, organizations can stay ahead in an increasingly competitive and technologically driven procurement landscape.

Private GenAI+RAG and scorecarding present a secure, efficient, and scalable RFP/RFI evaluation approach. Organizations can maintain compliance, enhance accuracy, and accelerate decision-making by keeping AI processing on-premises. As AI evolves, organizations should explore on-prem solutions to safeguard sensitive procurement data while leveraging the benefits of AI-driven automation.

References

[1] P. Rajpurkar, J. Zhang, K. Lopyrev and P. Liang, "SQuAD: 100,000+ questions for machine comprehension of text," 2016.

[2] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan and P. Dhariwal, "Language models are few-shot learners," 2020.

[3] S. D. Jap, "Online reverse auctions: Issues, themes, and prospects for the future," *Journal of the Academy of Marketing Science,* vol. 30, no. 4, pp. 506-525, 2002.

[4] P. D. Cousins, B. Lawson, and B. Squire, "Performance measurement in strategic buyer-supplier relationships: The mediating role of socialization mechanisms," *International Journal of Operations & Production Management,* vol. 28, no. 3, pp. 238-258, 2008.

[5] S. Talluri and J. Sarkis, "A model for performance monitoring of suppliers," *International Journal of Production Research,* vol. 40, no. 16, pp. 4257-4269, 2002.

[6] H. Touvron, T. Lavril and G. Izacard, "LLaMA: Open and Efficient Foundation Language Models," 2023.

[7] K. Guu, K. Lee and Z. Tung, "Retrieval augmented language model pre-training," in *International Conference on Machine Learning*, 2020.

[8] P. Lewis, E. Perez, and A. Piktus, "Retrieval-augmented generation for knowledge-intensive NLP tasks," 2020.

[9] S. Borgeaud, A. Mensch and J. Hoffmann, "Improving language models by retrieving from trillions of tokens," in *International Conference on Machine Learning*, 2022.